RESEARCH DESIGN FOR STUDY OF FACTORS RELATED TO PHYSICIAN PRICES

Pennsylvania Blue Shield

RESEARCH DESIGN FOR
STUDY OF FACTORS RELATED TO PHYSICIAN PRICES

Research Report 77-2

January 1977

Prepared for

Social Security Administration

Under Contract No. 600-76-0146

RESEARCH GROUP

PLANNING AND RESEARCH DIVISION

PENNSYLVANIA BLUE SHIELD

Camp Hill, Pennsylvania 17011

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

I. INTRODUCTION AND GENERAL APPROACH STRATEGY

This project seeks to develop knowledge and insight into the pricing practices of Pennsylvania physicians and into the relationship of physicians' fees to physician characteristics and to the characteristics of their practice localities. The studies emphasize use of source data from files of claims paid by Pennsylvania Blue Shield (PBS) both as Medicare Part B intermediary and as a private health insurance carrier throughout the state of Pennsylvania.

There is a generally acknowledged need for research to develop credible and defensible substantive data on prices of physician services. This should include a critical analysis and a fresh new appraisal of the reported charges in carriers' paid claims data bases if these data bases are to serve as a key source from which such information is to be derived. There also is a need for research to develop better knowledge and understanding of the factors related to, and of the processes driving, physician price behavior. Finally, there is a need for research pointing toward applications, involving the formulation and testing of alternative concepts for using the findings of basic research such as that just described in connection with the structuring of improved reimbursement mechanisms.

Distributions of reported charge data for individual physicians will be described and analyzed in this study with the intent of determining the most appropriate measures of representative price for a given service by a given physician. Critical examination will be made of various methodological alternatives for aggregating these physician charge data and obtaining measures of representative price (and price range) for given specialties and/or localities, as well as for all doctors within Pennsylvania Blue Shield's service area. These data will be used in statistical analyses, along with appropriate socioeconomic data and data

on the characteristics of individual physicians, to draw out and explore the principal relationships between physicians' fees and the characteristics of physicians and their practice localities. The main thrust of the analytical portions of this study will be to test a variety of hypotheses that would characterize physicians as price discriminators, and to identify and measure the relative impact of variables in terms of which such discrimination may be explained. Finally, analyses will be conducted to guide the selection and design of alternative locality, specialty, or other classifications that could be used in constructing prevailing fee screens for physician reimbursement under either Medicare or private UCR systems. These alternative classification systems for physician reimbursement will be compared and evaluated, primarily on the basis of comparative payout simulations. Results obtained from the Medicare and private business aspects of the study will be compared with each other as may be appropriate to identify any salient likenesses or differences in pricing practices under private insurance and Medicare.

Again, it is emphasized that such studies should begin with a thorough analysis of the source data, rather than simply use the usual, customary and prevailing charge data already in existence in carriers' UCR data bases and profile systems. While it is true that data exist on specific medical procedures for each locality and specialty classification employed by a carrier in its reimbursement programs, both for Medicare and private business UCR, it cannot necessarily be taken for granted that these data are uniformly accurate and valid.

The general strategy is to approach the study in three main phases, as already suggested above. The first phase is an essentially descriptive study of physicians' prices and price trends; the second is a study of the relationships between physicians' prices and various factors that may be used to

characterize individual physicians and their practice locales; the third phase is an exploration and comparative evaluation of alternative schemes that may be devised to segregate physicians into peer classes for reimbursement purposes. These three phases of study are discussed in three separate following sections of this report. First, however, follows a description and discussion of physician price data resources available at PBS for purposes of this research program.

II. PRICE DATA FROM PBS FILES

In the process of administrating regular business health insurance and various government programs, PBS receives and processes claims for expenses incurred by subscribers and beneficiaries. The specific services reported on each claim are recorded on one or more line items which describe the various procedures, identified by type-service and procedure codes, for which the doctor is billing. Information on each line item is coded and entered into the computerized claims processing system as a record on a computer data file. Initially these records contain only information supplied by the doctor such as his charge, the number and place of treatments of each procedure, and the patient's age and sex. In the course of processing claims, additional information is developed and coded on each line item, including the amount to be allowed and the final disposition of the claim (i.e., paid in full, rejected or partly paid). After the claim records have completed their journey through the claims processing system and finally have been paid or rejected, selected data items are extracted from the claim records and are saved for subsequent statistical and financial analysis.

DATA EDITING AND QUALITY CONTROL

Experience data on the above extract files are useful in analyzing trends in total utilization and certain financial statistics, but some of the claims reflect unusual clinical circumstances or contain coding errors which prohibit their inclusion in many kinds of research analysis. To eliminate such data, the Research Department edits the above extract files, screening out data that would be

undesirable for research purposes. The following are reasons for which data are deleted:

(1) Disposition indicates line item was not paid.

(2) Zero charge or zero allowance (payment) indicated in record.

(3) Medicare claims from hospital-based physicians. (Since payment of these claims is shared with and mostly assumed by the Part A carrier, Part B allowances on such claims are much lower than for other claims.)

(4) The number of services (treatments) indicated on a line item exceeds a predetermined limiting value, indicating either an abnormal circumstance or a likely coding error in the number of services.

(5) The claim contains an action code, indicating that unusual clinical circumstances were involved which could produce an unusually high charge.

These edits filter out between one and two percent of the data on the extract files, and improve the accuracy and reliability of the remaining data on the edited Research Paid Claim Files. This is particularly important when the data are being used for price studies. The unit charge made by a doctor for a procedure generally is calculated as his total charge in a line item for the procedure divided by the number of services indicated in the line item. If, for example, the number of services is incorrectly coded, the value read by the computer could be an order of magnitude above the true value. When such errors occur, their usual effect is to cause the unit charge to appear unrealistically low, as well as distorting the apparent utilization rate. The Research edits of services, charges, etc. which are designed to clean up the data, cost little in the way of lost data and greatly reduce the chance of errors such as those suggested above.

PROJECT DATA BASE

Regular business and Medicare paid claims experience have been collected on Research files since 1968. However, some of these files were created prior to the development of the above edit procedures, and as such, contain an unknown amount of data of questionable validity. Moreover, much of the data contained in these files would be either unnecessary or irrelevant to the specific purposes of this study. For reasons of economy and efficiency, special files are being extracted for use in this project. These files contain only data items thought to be relevant to the study and which also are known to be of good quality (i.e., reliably coded and maintained). The data items contained in the files generally describe the patient, the reported services, prices and PBS action.

In addition to the items already discussed, information on the specialty, residence, etc. of each doctor who has done business with PBS is maintained on the company's Vendor Master files. These items have been merged into the claims experience files on the basis of doctor identification numbers on the experience line items. Note should be made, however, of an important distinction between the data on physician characteristics and the data obtained from claim transactions. The data on physicians represent each doctor's status at a fixed point in time, frequently updated. Outdated versions of these files do not exist prior to March, 1976. Therefore, it is necessary to attach current information on physicians to claims which they submitted in prior years.

Specific data items extracted from PBS sources and maintained on special files created for this project are the following:

Doctor Number

County of Doctor's Residence

Doctor's Specialty

Doctor Type (M.D., D.O., D.P.M.; PBS Participation)

Coverage Code (Regular Business - Prevailing Fee or Fee Schedule/Medicare)

Type Service/Procedure Code

Billed Charge

Allowed Charge (not available for private business)

Payment (not available for Medicare)

Number of Services

Incurred Date of Service

Date of Claim Payment ("approval," for Medicare)

Place of Treatment (in-hospital, office, etc.)

Assigned Status (indicator of whether doctor accepted assignment on claim - Medicare only)

FILE ORGANIZATION

The PBS Regular Business and Medicare experience files are maintained separately, and experience data for each of these lines of business have been organized into a separate file for each quarter by date of claim payment. The data on these files represent claims paid for 1970 through 1976, with the exception of Medicare claims approved during the period September 1970 through December 1970, and during April 1971, for which experience files have been lost.

Also, although experience files have been collected for 1970 through 1976, some of this data may not be usable in the study due to changes over the years in procedure coding conventions. In order to perform longitudinal analyses of changes in physicians' charges, it is imperative that comparisons of charges across time periods be based on comparable procedure definitions. Translations are available which relate some procedure codes under the various coding systems. Substantial amounts of data nevertheless will inevitably have to be left out of the longitudinal investigations due to the incompleteness of accurate translations.

III. PRICE DATA DEVELOPMENT AND ANALYSIS

The initial phase of the project is an in-depth analysis and evaluation of the physician charge data contained in PBS paid claim files for Medicare Part B, and for both the UCR and Fee Schedule lines of private insurance business.

PBS offers and operates two distinct types of programs as part of its regular private business -- Fee Schedule Programs and UCR (Prevailing Fee Programs). Physicians practicing in Pennsylvania also have the choice of being, or not being, "Participating Physicians" with PBS. Being a "Participating Physician" obligates the physician to accept, as payment in full, PBS's determination of reasonable price under the UCR program. "Participating Physician" for PBS Regular Business is equivalent to a physician who always "accepts assignment" under Medicare. It is intended that physician charges reported under both the Fee Schedule and UCR private programs will be utilized in these studies as well as Medicare data.

The objectives of this initial phase of study may be summarized as follows:

(1) Using PBS paid claim experience files as a source data base, describe and investigate the characteristics of physician charge data distributions for:

   (a) individual doctors;

   (b) given physician specialties;

   (c) given localities; and

   (d) multiple-specialty and multiple-locality aggregates.

(2) Determine the best ways of obtaining and expressing measures of representative price for selected physician services at the level of:

(a) individual doctors;

(b) given specialties;

(c) given localities; and

(d) multiple-specialty and multiple-locality aggregates.

(3) Compile substantive data on representative prices in the state of Pennsylvania for a selected list of benchmark procedures, with breakdown by specialty and locality to the extent that is practical and feasible.

(4) Test the hypothesis that individual physicians are price discriminators; and where discrimination is detected, see if it tends to be associated predominantly with certain specific procedures or identifiable situations.

(5) Test the hypothesis that there is price discrimination among specialties and/or between different geographic locales.

(6) Test the hypothesis that physician fees are tending to become less disperse over time.

PRICE DISCRIMINATION

Some of the questions of greatest interest are those relating to the possibility of price discrimination. There are a number of possible tests that could be applied in seeking answers to these questions. One of these would be to test for the existence of multiple modes in the distributions of physicians' charges. When multimodal cases are in fact found, however, they must be interpreted very carefully, since they could be indicative of any of the following conditions: (a) it could be a time related phenomenon, representing the effect of a price change during the time interval spanned by the data; (b) it could be indicative of an ambiguously

defined procedure; (c) it could be a reflection of some effect induced by differences in reimbursement methods; (d) it could be a reflection of something about the physicians themselves, variations in practice localities, or in the populations served. Whenever a multimodal distribution is detected, an early attempt will be made in this phase of study to establish whether it can be explained simply as the result of a price change or an ambiguously defined procedure. If not, it generally should be safe to assume that it represents price discrimination arising from one of the latter two classes of conditions ("c" or "d") listed above. Further investigation in subsequent stages of study could confirm this and provide some understanding of the nature of the discrimination.

A second -- perhaps easier -- type of test would be to compare various distributions of charges by examining the relationship of the mean to the standard deviation in each distribution. The larger the standard deviation relative to the mean, the more likely it is that some discrimination has occurred. One also could test for discrimination by calculating and comparing coefficients of variation. The hypothesis would be that physicians or procedures with relatively higher coefficient of variation may be an indication of discrimination. Still another approach could be an examination and comparison of measures of Kurtosis for various distributions of charges. Kurtosis alone probably does not provide a sufficient test, however, since it is somewhat unclear whether discrimination would be more indicated by "fat" tails (high Kurtosis) or by broad and flat distributions (low Kurtosis). Kurtosis measures most likely will be of value for these purposes only when considered in conjunction with other measures and indicators.

BENCHMARK PROCEDURES

The analyses to be conducted in this study will be addressed in terms of a

group of selected benchmark procedures. The following general criteria apply to the selection of benchmark procedures:

(1) The definition and description of the procedures must be widely understood and generally accepted. It should conform to conventionally accepted patterns of provider practice.

(2) The procedure should occur frequently enough to provide adequate sample sizes and afford an opportunity for thorough analysis.

(3) Where possible, the procedure should be reported across a reasonably large number of specialties. However, some procedures will be included which are usually performed almost exclusively by a single specialty.

(4) The benchmark procedures should be selected to be reasonably representative of the actual mixes of services reported to PBS.

Selection of benchmark procedures which are well-defined and generally accepted can be accomplished by concentrating on procedures which have been included in the benefit structure for a long enough period (three to five years) so as to become widely accepted and recognized by providers. Procedures which are relatively new, and/or on which there may be substantial variation in the complexity or difficulty of the service actually performed from one case to another, are routinely classified as procedures requiring individual consideration (IC); and claims for such services then are suspended from the normal claims processing system for manual pricing. Procedures normally treated as "IC" are to be excluded from the study.

Criteria concerned with the frequency and distribution of procedures will be addressed by considering procedure codes within the framework of PBS type-service categories. The type-service categories employed at PBS are generic classifications of provider activity such as surgery, in-hospital medical care, radiology,

etc. While type-service categories may tend to reflect groupings of procedures germane to a particular provider specialty, "type service" is not necessarily specialty oriented; e.g., surgery is performed by providers other than surgeons, and in-hospital medical care is provided by a majority of providers regardless of specialty.

The following is a list of 44 procedures which can be considered candidates for inclusion in the PBS private business benchmark set. Except for editing to exclude certain procedures already known to be poorly defined, or to eliminate certain procedures extremely similar in content to others already in the list, these are the most frequently reported procedures in each of nine major type-service categories. The 44 procedures listed below account for approximately one-half of the reported services and one-third of all payments made under Pennsylvania Blue Shield's private insurance business in calendar year 1975.

Type Service	PBS Procedure Code and Terminology	% Total Services	% Total Payments
ORAL SURGERY	07240 - Surgical extraction of one (1) tooth, complete bony impaction	.10	.30
	07243 - Surgical extraction of four (4) teeth, complete bony impaction	.08	.72
	07230 - Surgical extraction of one (1) tooth, partial bony impaction	.08	.13
SURGERY	20610 - Arthrocentesis, major joint	1.19	.53
	45300 - Proctosigmoidoscopy	1.00	.69
	11460 - Excision of benign lesion, diameter greater than 1.8 cm or complicated	.43	.78
	53660 - Dilation urethra	.43	.14
	11480 - Excision of lipoma - subcutaneous or sebaceous cyst, over 1" and up to 3"	.31	.40

Type Service	PBS Procedure Code and Terminology	% Total Services	% Total Payments
SURGERY (Continued)	12020 - Simple repair of a recent wound, up to 2.5 cm, in face or genitalia	.26	.15
	17210 - Plantar verruca, removal by any method except excision, not exceeding 13 mm	.25	.12
	58120 - Dilation and Curettage (D&C)	.21	.64
	52000 - Cystourethroscopy, in office	.19	.24
	58150 - Hysterectomy, total	.16	2.11
	17000 - Electrosurgical destruction of lesion, single	.16	.10
	49500 - Herniorrhaphy, inguinal, unilateral	.12	.89
	19120 - Excision breast cyst, unilateral	.12	.38
	47600 - Cholecystectomy	.10	1.20
	55250 - Vasectomy	.10	.40
	42840 - Tonsillectomy, child	.10	.36
OBSTETRICS	59400 - Delivery, vaginal, including antepartum and postpartum care	.57	4.64
ANESTHESIA	59410 - Anesthesia for obstetrical delivery	.13	.20
	58120 - Anesthesia for D&C	.09	.12
	58150 - Anesthesia for total hysterectomy	.08	.26
	49500 - Anesthesia for herniorrhaphy	.06	.12
	47600 - Anesthesia for cholecystectomy	.05	.17
DISGNOSTIC X-RAY	71020 - Chest, PA and lateral	1.67	1.08
	71010 - Chest, single view	.44	.18
	74240 - Upper GI	.43	.55
	74290 - Cholecystography	.24	.21
	73570 - Knee, 3 or 4 views	.21	.13

Type Service	PBS Procedure Code and Terminology		% Total Services	% Total Payments
IN-HOSPITAL MEDICAL	90260-	Hospital visit, intermediate	12.59	3.89
	90240-	Hospital visit, brief	7.50	1.86
	90270-	Hospital visit, extended	6.47	2.64
	90220-	Hospital visit, initial, with comprehensive history and physical exam	2.71	2.99
DIAGNOSTIC MEDICAL	93000-	Electrocardiogram with interpretation and report	4.05	2.34
	95820-	Electroencephalogram	.11	.14
	93260-	Cardiovascular stress testing	.09	.20
PATHOLOGY	85010-	Complete blood count (CBC)	1.50	.28
	84330-	FBS--Sugar glucose blood	1.07	.15
	81000-	Urinanalysis routine, complete	1.04	.11
	82465-	Cholesterol blood	.54	.10
	84475-	Triglycerides blood	.41	.11
CONSULTATION	90620-	Consultation, comprehensive	1.54	1.83

While many of these same procedures also occur frequently in Medicare experience, the high-frequency set for Medicare is in large measure differently constituted. The lack of conformity in composition of the two sets is a reflection of differences in both basic benefit structure and age of population served. The following is a list of 33 candidate procedures for the Medicare benchmark set selected on the basis of being the most frequently reported procedures in each of nine major type-service categories for Medicare Part B. These 33 procedures account for 71 percent of the reported services and 54 percent of the allowed charges for Medicare Part B in Pennsylvania Blue Shield's service area for calendar 1975.

Type Service	PBS Procedure Code and Terminology	% Total Services	% Total Payments
SURGERY	11701 - Debridement of mycotoc nails, multiple or complicated	.54	.53
	20610 - Arthrocentesis, major joint	.47	.39
	45300 - Proctosigmoidoscopy	.25	.35
	11720 - Avulsion and/or excision of nail, partial, single	.16	.14
	11010 - Debridement of ulcer, initial	.13	.10
	66900 - Extraction of lens unilateral, for cataract, dislocated lens, etc.	.09	2.71
	52000 - Cystourethroscopy, in office	.08	1.89
	52600 - TUR Resection prostate, transurethral, complete	.06	1.89
ANESTHESIA	79502 - Anesthesia for computerized transaxial brain scan	.35	.65
	66900 - Anesthesia for extraction of lens unilateral, for cataract, etc.	.04	.22
	52600 - Anesthesia for TUR Resection prostate	.03	.16
DIAGNOSTIC X-RAY	71020 - Chest, PA and lateral	.59	.68
	71010 - Chest, single view	.31	.22
	74240 - Upper GI	.10	.24
	79500 - Computerized transaxial brain scan	.09	.11
	74270 - Colon, single procedure, routine barium enema	.07	.14
OFFICE MEDICAL	90020 - Office visit, intermediate	11.71	7.49
	90005 - Office visit, brief	11.22	5.09
	90030 - Office visit, extended	4.96	3.67
	90000 - Office visit, with injection, etc.	1.61	.28

Type Service	PBS Procedure Code and Terminology	% Total Services	% Total Payment
OFFICE MEDICAL (CONT'D)	90040 - Office visit, comprehensive history and physical exam	.65	1.02
HOME MEDICAL	90110 - Home visit, brief	1.36	.72
	90120 - Home visit, intermediate	1.35	.93
	90130 - Home visit, extended	.90	.71
IN-HOSPITAL MEDICAL	90260 - Hospital visit, intermediate	12.30	7.74
	90240 - Hospital visit, brief	6.55	2.95
	90270 - Hospital visit, extended	4.56	3.60
	90220 - Hospital visit, initial, with comprehensive history and physical exam	1.61	3.38
DIAGNOSTIC MEDICAL	93000 - Electrocardiogram with interpretation and report	1.77	2.11
	93061 - Transtelephonic pacemaker analysis	.36	.57
	95820 - Electroencephalogram	.13	.17
PATHOLOGY	84330 - FBS - Sugar Glucose blood	2.05	.54
	81000 - Urinanalysis routine, complete	1.64	.30
	85010 - Complete blood count (CBC)	1.16	.42
	85610 - Prothrombin Time	.58	.16
	84520 - BUN--Urea-nitrogen blood	.36	.11
	82465 - Cholesterol blood	.35	.12
CONSULTATION	90620 - Consultation, comprehensive	1.11	2.66

The preceding lists are intended simply to identify some of the specific procedures reported most frequently under PBS private business and Medicare, respectively, and are not to be construed as a final proposal on how the benchmark sets should be constituted. The exact composition of the benchmark set(s) to be used in these studies will be determined only after further discussion and agreement between PBS and SSA(ORS). It is as yet unclear, for example, whether it would be more desirable to develop a single benchmark set to span both Medicare and PBS private business, or to develop separate benchmark sets acknowledging and reflecting the natural differences between these two distinctly different lines of business. Perhaps the best course would be to develop separate, but very similar, benchmark sets.

In finalizing the benchmark sets, effort should be made to include a number of procedures common to both Medicare and private business. Procedures also should be included which, although not necessarily among those most frequently reported, account for a relatively high proportion of payout. Finally, consideration also can be given to including the procedures used for the physician fee component of the Consumer Price Index. In the surgery category, for instance, these guidelines could result in a basic benchmark list such as the following:

Hysterectomy, total

Herniorrhaphy

Cholecystectomy

Tonsillectomy

Extraction cataract

TUR prostate

Appendectomy

Arthrocentesis, major joint

Proctosigmoidoscopy

DESCRIPTIVE STATISTICS FOR INDIVIDUAL PHYSICIANS

Initial data manipulations, as described in this subsection, will be performed in a strictly cross-sectional framework. Data for this stage of study will be taken from PBS claims experience spanning calendar year 1975. In general, there is no intention to merge the PBS private business data base and the Medicare data base for any of these studies. Rather, separate and parallel lines of investigation will be pursued for these two major lines of business, with cross comparisons being made to whatever extent may be found feasible and appropriate. The following discussion applies to each of these lines of inquiry.

The first step is to sort the 1975 claims data base by procedure and by individual doctor. For each benchmark procedure included in the study, a distribution will be generated indicating the relative numbers of doctors reporting with given frequencies. This can be displayed in both graphic and tabular form, and will assist in reaching a decision on the minimum number of services a doctor must have reported for a given procedure in order to have his data for that procedure included in any further analysis. In keeping with conventional practice, the aim will be to include a given doctor's data for a given procedure as being statistically meaningful only if he has reported that procedure at least thirty times during the covered time interval. There are, on the other hand, a great many procedures for which very few individual doctors will have reported this many claims during the course of any given year. It is conceivable that this could be found to be true for some of the selected benchmark procedures, even though they will have been selected partly on the basis of their frequency of occurrence. The data distributions generated at this stage will allow such situations to be detected, and if necessary, provide guidance to assist in establishing a compromise (i.e., smaller) minimum threshold frequency for certain procedures.

For each doctor-procedure combination satisfying the minimum frequency criterion, a further preliminary step is to determine whether the doctor has changed his price for that procedure during the time period spanned by the data base. A simple but effective way to test for this will be to look at the data by quarters, compare the doctor's mean charge for the fourth quarter to his mean charge for the first quarter of the year, and use one of the standard tests (e.g., t-test) for significance of difference. If the sample sizes for only a quarter of experience are too small to justify the conventional normal parametric tests, a nonparametric test such as Kruskal-Wallis still could be applied quite validly. The assumption would be that a significant time-related difference here probably represents a formal price change. Once a price change has been detected, there are a variety of options on what to do about it: (a) completely discard the doctor-procedure combination for which the change occurred, and suffer the data loss; (b) ignore it and simply still use all the data, suffering the effects of the perturbation in the data distribution; (c) compromise by determining the point in time at which the change occurred and then either using only data associated with the original price level, data associated with the latest price level, or data for the price level that was in effect for the greater part of the year. In fact, any one of these options could be invoked in any given instance, since a preference for any particular option should rightly depend upon the question being addressed, the requirements of the specific problem at hand, and the special circumstances that may obtain in a particular instance. For example, in attempting to understand the general characteristics and mechanics of these data distributions, it probably would be best to deal only with data for the price level that can be identified as having been in effect for the greatest part of the year; but for longitudinal studies, in which one may be dealing with time series of average prices for various component time periods over an extended

period of time, it probably would be better to use all the data for each component time period. In any event, it is important to be aware of the fact that price changes may have occurred in certain cells of the data structure, so that appropriate steps can be taken or results interpreted accordingly.

After the preliminary analyses discussed above have been conducted, the data base can be refined to include only those doctor-procedure combinations meeting the minimum requirements for number of services reported (i.e., satisfying at least minimal sample size requirements), and also incorporating whatever steps may be felt necessary and desirable in light of any individual price changes that are detected. The immediate next step should be to examine and classify each doctor-procedure data distribution as either unimodal, bimodal, or multimodal (defined here to mean three or more modes). Even though there will be a very large number of distributions to be so examined and classified, this can be readily accomplished by computer-aided methods. A summary listing will be generated, indicating for each procedure the relative numbers of reporting physicians whose charge data distributions had one mode, two modes, or three or more modes. If there are any procedures for which individual doctors predominantly tend to exhibit more than one mode, the definitions and content of these procedures will be carefully reviewed to see if there is anything inherent in the nature of the procedure itself that might explain the bimodal or multimodal distribution. As suggested earlier, in the discussion of tests for price discrimination, any biomodal or multimodal situations that at this stage cannot be suitably explained in terms of something about the procedure itself probably represent price discriminating situations.

The following statistics will be generated to provide a summary characterization of the distribution of reported charges for each doctor for each procedure:

(a) measures of central tendency -- mean, median and mode;

(b) measures of dispersion -- range, semi-quartile range, variance and standard deviation;

(c) skewness measures;

(d) Kurtosis measures.

One of the study objectives is to seek some generalization on which one of the conventional measures of central tendency provides the best single-valued representative of an individual doctor's distribution of charges for a given procedure. The "best" measure here will be defined as that one which yields the smallest mean deviation when compared to all values in the interquartile region of the distribution of charges. Mean deviation is the mean distance of the n sample members from the average value, where all distances are taken as positive, and "average" may be interpreted to be any measure of central tendency. In the present case:

$$MD = \frac{\sum_i \left| \bar{X} - x_i \right|}{n}$$

where $\bar{X}$ may be either the mean, the median or the mode, and the mean deviation (MD) is to be calculated for each of these three cases; and where the x_i are the n actual charge values contained in the region spanning the second and third quartiles of the charge distribution. The mean deviation here is defined and calculated only over the interquartile range, instead of over the entire distribution, because it is assumed that the data of primary interest in determining a "typical" value should be concentrated in the middle of the distribution, and the tails of the distribution may contain some highly atypical values. A summary listing will be generated for each procedure, indicating the relative number of doctors reporting that

procedure for which the mean provided the best single-valued representative price, the number for which the median was best, and the number for which mode was best. In evaluating these results to decide which measure of central tendency is "best," it should be helpful to supplement this information with some additional information on the distributions of individual doctors' means, medians and modes over all doctors for each procedure.

The distribution of individual doctors' variances, skewness and Kurtosis measures over all doctors also will be generated for each procedure. These outputs will be examined to identify those specific procedures that appear to be most subject to greater amounts of variance. It may be recalled that larger variances, or larger coefficients of variation, have been suggested as indicative of price discrimination. A thorough examination of the extent and patterns of variance itself varying from doctor to doctor, locality to locality, or procedure to procedure may provide some interesting new insights into the question of discrimination.

Findings of all the data development and analyses discussed above will be drawn together at their completion to produce a descriptive summary of the statistical characteristics of individual physicians' charge distributions for the selected benchmark procedures.

DESCRIPTIVE STATISTICS FOR AGGREGATES OF PHYSICIANS

The work described up to this point has concentrated on individual physicians. This then will be extended into development and statistical analysis of charge (price) data distributions for aggregates of physicians by specialty and/or locality of practice.

A variety of aggregation rules are available for use in developing data

distributions to characterize the pricing practices of collections of physicians. The usually preferred approach is to form a weighted aggregate of the statistical characteristics (e.g., average prices) for each doctor included in the collection. I.e., each individual doctor would be represented by his own average price for a given procedure, weighted by the number of services he reported for that procedure on claims submitted to PBS for payment. The principal alternative would be to use a simple aggregate of individual physicians' statistics, in which each physician's average value is represented once and only once, with no weighting to represent his volume of utilization. The difference between these two can be illustrated by observing that a payment screen established at the 75th percentile of the weighted distribution will isolate and eliminate the highest 25 percent of charges, whereas a screen at the 75th percentile of the unweighted distribution will isolate the highest charging 25 percent of all doctors. Either could be best for any given application, depending how one's objectives have been defined. A further alternative would be to aggregate all of the individual charge data transactions for all of the individual doctors included in the collection. This is the true distribution that the weighted aggregate of individual doctors' averages attempts to approximate. In a procedural sense, however, the true distribution generally is more unwieldy to develop and work with. While the weighted aggregate probably will be used for most purposes in this study, characteristics of the distributions formed by all three of these approaches will be examined and compared.

The derived data base of principal interest at this point will be a file containing each individual physician's average charge for each benchmark procedure for which minimum sample size requirements have been satisfied. The records in this file also will identify the physicians' specialties and the counties in which

they practice. This file will be used to develop distributions of the average charges for each specialty for each procedure, and also for each county for each procedure. Statistical parameters of these distributions will be developed and analyzed in a manner analogous to that described earlier for individual physicians' data distributions. Parameters developed will include various measures of central tendency, measures of dispersion, skewness and Kurtosis. Significance of differences in means will be tested; variances and Kurtosis measures will be evaluated for each distribution; bimodal and multimodal distributions will be identified and isolated. The hypotheses being tested will address the question of price discrimination both within and between specialties, and within and between geographic localities.

Results of this stage of the project work will be presented in a report summarizing and interpreting the statistical characteristics of average charge distributions for aggregates of physicians by specialty and geographic area.

ANALYSIS OF PRICE TRENDS

To examine the behavior of physicians' prices over time, longitudinal data will be extracted and organized to create a file of physicians' reported charge data for selected benchmark procedures. A given physician's data for a given procedure will be included if and only if he has reported the procedure often enough to satisfy the specified minimum required sample size.

The procedures to be included in longitudinal analyses will be limited to a relatively small subset of the complete benchmark set used for the 1975 cross-sectional studies. This is first and foremost a matter of technical necessity, since good translations do not exist to relate many of the procedure codes in the 1975 benchmark set to correctly comparable codes in the procedure coding and

terminology systems that were in use at PBS during certain earlier periods. The selected benchmark set for longitudinal studies will contain between ten and fifteen procedures, satisfying the following criteria: (a) they must have "clean" one-to-one translations between the various procedure coding systems that have been in effect at PBS since 1970; (b) they should be frequently occurring; (c) as many as possible should occur commonly in both Medicare and PBS private business experience; and (d) they should be selected to provide a reasonably representative balance between various major type-service categories. The following, for example, might comprise a very representative set of procedures satisfying these criteria:

Type Service	PBS Code	Procedure
SURGERY	45300	Proctosigmoidoscopy
	58120	D&C
	58150	Total Hysterectomy
	47600	Cholecystectomy
	42840	Tonsillectomy
OBSTETRICS	59400	Vaginal Delivery, including pre- and post-partum care
ANESTHESIA	59410	For obstectrical delivery
	58150	For total hysterectomy
DIAGNOSTIC X-RAY	71020	Chest, PA and lateral
	74240	Upper GI
MEDICAL	90260	Intermediate hospital visit
	90005	Brief office visit (Medicare only)
DIAGNOSTIC MEDICAL	93000	Electrocardiogram (EKG)
PATHOLOGY	85010	CBC - complete blood count
	81000	UA - urinanalysis

For the selected subset of procedures, time series data will be developed spanning the period 1970 through 1976 in six-month intervals. Statistical measures of central tendency, dispersion, skewness and Kurtosis will be determined for the data distributions in each of these six-month periods. These distributions and statistical measures will be developed for individual procedures, for type-service categories, individual specialties and geographic areas, and for the total experience spanned by the selected procedures.

Price movements throughout this seven-year period will be described in terms of time series of average prices and price relatives. This will require construction of a physician fee index tailored to the specific needs and dimensions of this project. Essentially, it involves the aggregation of weighted average prices for a representative mix of services in each time period, relative to a base period. This can be a complex and challenging problem. However, PBS has already developed computational methods and computer software to periodically generate physician price index data, which could be simplified and emulated on a smaller scale for purposes of this project.

Trend data also will be developed and examined to determine whether physicians' pricing practices are growing more uniform or more disparate. To test the hypothesis that the distribution of physicians' prices is becoming tighter over time, one can examine the time path in: (a) the average measure of Kurtosis over procedures; (b) the average relative coefficient of variation over time; and (c) the average proportion of charges that fall within 50 percent of the median charge.

PRICES UNDER DIFFERING PAYMENT METHODS

Pennsylvania Blue Shield offers two main kinds of private insurance programs: Prevailing Fee (UCR), which is essentially a full payment program for all covered

services; and Fee Schedule programs, which will never pay more than a prescheduled amount, usually leaving the patient/subscriber with a not insignificant out-of-pocket expense to cover the remainder of the doctor's total bill. The Plan B fee schedule, originally introduced in 1954, today tends to pay about half -- and frequently less than half -- of a doctor's charge for most kinds of services. The Plan A fee schedule, first introduced in 1945 and now almost completely phased out, tends to pay less than one-third. The Plan C fee schedule, which on the average tends to pay a little over two-thirds, was introduced to the market only in the second half of 1975 and comprised an almost negligible part of PBS private insurance business for that year.

Folk wisdom in the health insurance business has it that doctors' charging practices tend to reflect these differentials in payment plans. This may have been true at one time, and still may be true at some places or under certain circumstances. PBS studies in recent years, however, suggest that Pennsylvania physicians' charging practices do not vary significantly on the basis of the patient's insurance plan and implicit cost sharing responsibility. This continues, nevertheless, to be a question of great interest and will be investigated as one of the studies to be conducted under this project. The hypothesis to be tested is that doctors tend to charge higher prices to patients with more generous and complete payment plans. This serves as a surrogate for, and in effect may also provide a test of, the hypothesis that physicians charge higher prices to patients who have less cost sharing (or conversely, that they charge less to patients who have more cost sharing).

A straightforward and simple initial test of this hypothesis is to tabulate the average reported charges for each of the benchmark procedures under UCR and the various Fee Schedule plans respectively. Visual inspection of this tabulation

will provide some early insight into the patterns of variation in average charges reported under these various payment plans. Further analysis should concentrate primarily on comparisons between charging patterns for UCR and Plan B only, since Plan A and Plan C combined accounted for less than two percent of PBS private insurance business volume in 1975, and this experience may be too sparse to provide any meaningful results. A simple sign test between UCR and Plan B average charges will reveal whether higher charges tend usually to be associated with UCR. Previous tests of this sort have indicated UCR and Plan B each being higher about half of the time (i.e., each higher for about half of the procedures included in the test), which clearly suggests no significant difference. Repetitions of this test using data only for selected individual specialties or geographic areas may, however, begin to show some apparently significant differences within given specialties or localities. Previous PBS studies have not examined the question at this level.

A more powerful test of the hypothesis can be accomplished by regression analysis, in which individual doctors' average charges for a given procedure are regressed against the relative volume (percent) of each doctor's services reported under UCR. Analysis of the correlation coefficients from these regressions will reveal whether there is any significant relationship between average charges and cost sharing (as represented by payment mechanism). The hypothesis would be that higher charging levels are associated with higher proportions of UCR -- i.e., that there is a significant positive regression coefficient for average charges as a function of relative volume reported under UCR.

This analysis can be extended further to examine the possibility of differences in pricing practices between Medicare and PBS private business, as well as between the UCR and fee schedule facets of the private business only. For this analysis, a sample of doctors would be selected who have

reported experience under all three lines of business -- i.e., Medicare, private UCR, and private fee schedule. For some small number of representative procedures, the average prices charged by each individual doctor under each of these three lines of business would be calculated, the significance of any observed differences tested, and the results summarized and displayed in tabular and/or graphic form.

IV. ANALYSIS OF FACTORS RELATED TO PRICE

Analyses will be conducted using conventional multivariate statistical methods to identify those physician characteristics and socioeconomic characteristics of practice localities that appear to be most saliently related to physicians' price levels. These investigations will analyze, describe and demonstrate variations in price levels corresponding to variations in key parameters most strongly correlated with price.

Analyses also will be conducted to explore the potential for partitioning the physician population into discrete groups representing distinctly different empirical levels of pricing practices -- i.e., to segregate "high chargers" from "low chargers" in various gradations. Further analysis then will seek to identify those variables -- physician and locality descriptors -- in terms of which this partitioning can be explained, and to identify ranges of each explanatory variable associated with each partition in the overall range of prices.

These will be cross-sectional studies, using the data base of average prices for individual physicians from PBS claims experience for 1975, together with physician data and socioeconomic data available for whatever time periods most closely coincide with the 1975 time frame.

SOCIOECONOMIC DATA

Data on the socioeconomic characteristics of localities will be obtained from standard reference sources, principally from the 1976 edition of the Pennsylvania Abstract (Pennsylvania Department of Commerce) and from General Social and Economic Characteristics, Pennsylvania, 1970 Census of Population

(U. S. Department of Commerce). Additional data for counties and municipalities are available from the 1976 editions of the Pennsylvania Industrial Census Series published by the Pennsylvania Department of Commerce. Most of the socioeconomic data used in these studies can be expected to come from the 1970 census, with data from interim updates of the 1970 census being used where ever available.

The basic catchment area for data to be used in these studies is at the county level. Following is a list of those data items readily available by county that would be potentially useful for purposes of these studies.

General and Social Characteristics of Counties

- Population density
- Percent urban vs. rural
- Percent of population aged 65 or older
- Education (probably only one of the following):
  - Median school years completed
  - Percent high school graduates in population
  - Percent college educated
- Housing: percent of units owner occupied

Economic Characteristics

- Income level of population (probably one of the following):
  - Mean income
  - Median income
  - Per capita income
- Unemployment
- Occupational characteristics:
  - Percent professional, managerial, etc.
  - Percent craftsmen, foremen, etc.

Percent operatives

Percent laborers

Percent farm-related

Median value of housing units

Welfare statistics (probably one or two of the following):

Percent of families under poverty level

Percent of population receiving public assistance

Per capita expenditures for public assistance

Health Related Resources and Activity

Number of hospital beds

Beds in homes for aged, infirm, etc.

Physician density, relative to population

Percent of employed persons working for hospitals and health services

Public expenditures by state sources for medical care

Volume (or per capita volume) of Medicare Part B payments

Volume (or per capita volume) of Blue Shield payments

Percent of doctors participating with Blue Shield

Medicare assignment acceptance (one or more of the following):

Percent of Medicare Part B services on which assignment is accepted

Percent of doctors accepting assignment for at least x percent (80%?) of services delivered

Percent of doctors accepting assignment for x percent (20%?) or less of services delivered

Percent of Blue Shield services paid under UCR

It may be noted that some of the items in this list are apparently redundant. Mean income, median income and per capita income, for example, are alternate ways

of measuring what is essentially the same thing, and as such, would not be an independent subset of input variables. Preliminary analysis will attempt to determine which of these relates most strongly to prices of physician services in an area, and in general that can be expected to be the measure used for all further analyses. Other dependent subsets will be similarly analyzed.

Physicians' practice localities can be pinpointed somewhat more narrowly by defining sub-areas within counties in terms of the ZIP code zones they encompass, and then identifying a given physician with a given area on the basis of the ZIP code in his mailing address. This would allow physicians' practice localities to be identified, for instance, as being in either the inner city or urban fringe areas of each of Pennsylvania's several metropolitan areas. Most of the socio-economic data items, however, would be quite difficult -- if not impossible -- to isolate at these finer levels of resolution. The possibilities for accomplishing this, nevertheless, will be investigated, since some interesting additional questions could be explored if it is found possible to work at this level of resolution. E.g.:

(a) Are the major metorpolitan areas similar to each other in terms of pricing practices and/or socioeconomic characteristics?

(b) Are lesser metropolitan areas similar to, or different from, the major metropolitan areas?

(c) Are the lesser metroplitan areas similar to each other?

(d) Are the inner city areas different from the urban fringe areas?

Data also can be obtained and organized by Standard Metropolitan Statistical Area (SMSA), if so desired. Since half of Pennsylvania's twelve SMSA's are multiple county areas, this would provide a somewhat less fine level of resolution than data organized strictly by county. At present no studies are contemplated which would

use data organized by SMSA. However, it would be possible to shift to this basis for certain aspects of the studies if it is found necessary or expedient to do so.

CHARACTERISTICS OF PHYSICIANS

Data describing the characteristics of individual physicians will be taken primarily from the 26th Edition of American Medical Directory (AMA, 1974) and from PBS files. Data items available from these sources that may be useful for purposes of these studies are the following.

From American Medical Directory

- AMA membership: yes/no
- Physician's age
- Medical school attended
- Years since M.D.
- AMA specialty of record
- Board certification

From Blue Shield (PBS) Sources

- Specialty of record at PBS (self-designated)
- Type of practice: solo, partnership, group
- Blue Shield participation: yes/no
- Location of practice (county)
- Percent of Medicare Part B services on which assignment accepted (1975)
- Total Blue Shield payments received (1975)
- Total Medicare Part B payments received (1975)
- Percent of Blue Shield services paid under UCR (1975)

Data contained in the 26th Edition of American Medical Directory are for the year 1973. This is the most recent edition available. Data for the first four items on the above list for Blue Shield sources will come from the PBS Vendor Master File and will reflect the status of each individual physician in that file as of March, 1976. This is the oldest version of that file still available; but it should be quite compatible for matching with 1975 experience data, since changes occur very slowly in the Vendor Master File. The last four items in the above Blue Shield source list will be derived from PBS paid claim experience files for 1975.

PRELIMINARY ANALYSIS OF VARIABLES

The principal analytical tools to be used in studying the relationship of the above variables to prices of physician services are multiple regression and discriminant analysis. Before beginning these multivariate analyses, however, it would be wise to undertake some preliminary analysis and evaluation of the interactions between these various non-price variables, and also to gain some understanding of the simple bivariate relationship of each individual variable with price. The purposes of this stage of investigation are: (a) to identify sets of non-price variables that are strongly correlated, and which therefore would fail to adequately satisfy the independence requirements of the multivariate models; (b) to provide information to help determine the best variables to retain, and those most appropriately eliminated, from dependent subsets of non-price variables; and (c) to provide an early detection of those individual variables to which price is apparently strongly related, those which bear no apparent relationship whatsoever to price, and those which may have a markedly nonlinear relationship with price. In sum, this part of the study will provide valuable insight and

understanding to help guide the formulation of specific models and hypotheses for the multivariate studies.

Pair-wise independence of the physician descriptor variables can easily be tested by use of contingency tables. This is appropriate particularly when some of the variables are qualitative, as is the case for some of the physician descriptors. For example, physicians can be classified both by their specialty and by the county in which they practice. If these two classifications (i.e., descriptor variables) are independent, the expected representation of different specialties in each county should be proportional to the relative numbers of doctors in each of these specialties in the state as a whole; and similarly the distribution of each specialty across counties should be proportional to the relative numbers of all doctors in each county. The contingency table analysis counts the number of doctors actually assignable to each cell of the cross-classification matrix, and uses a chi-square test to determine the significance of the difference between these actual frequencies and the expected frequencies. The example just given would be testing the null hypothesis that there is no relationship between physicians' choices of specialty and location of practice. We might anticipate in this case that a significant relationship will in fact be found. Other pair-wise combinations of physician descriptor variables can be similarly tested, with similar statements of hypotheses.

To test the pair-wise independence of socioeconomic variables, the contingency table analysis would not be appropriate. Cross-classification of only 67 counties where there are potentially several different levels of each classification parameter would inevitably produce a number of low frequency cells in the cross classification, which would cause the analysis by this approach to break down rather badly. For this reason pair-wise relationships between the socioeconomic variables

are better investigated by simple bivariate regression analysis. Median income, for instance, could be regressed on population density to test the hypothesis that median income varies as a function of population density -- or more specifically, to test the hypothesis that median income increases as population density becomes higher. Examination of the correlation coefficient from this regression, moreover, would test the hypothesis that these two variables are significantly related. Strictly speaking, in the extension of this to other pair-wise combinations of socioeconomic variables, one would be testing a series of null hypotheses of the form H_o: $r_{ij}=0$, where the r_{ij} are the correlation coefficients of variable i with variable j for all combinations of variables i and j to be so tested.

Pair-wise tests of independence and/or functional relationship between individual socioeconomic variables and individual physician descriptor variables (i.e., a socioeconomic variable with a physician descriptor variable) would be accomplished by a combination of bivariate regression analysis and contingency table methods. The selection of method to be used in any given test would depend upon an examination of the nature of the specific two variables involved. In testing for interactions between socioeconomic variables and physician descriptor variables, the test series would not necessarily exhaustively test every possible pair-wise combination between variables of these two main types. There is no a priori common sense reason, for example, to expect that there would be a relation between the relative number of board certified doctors in a county and the percent of that county's population aged 65 or older.

The above series of pair-wise tests for independence or functional relationship between non-price variables will be reinforced by a cluster analysis on variables. The cluster analysis on variables takes as input the observed values of each variable for each subject (locality or physician). It looks at all variables simulatneously and sorts them into sets, or clusters, or related variables

on the basis of some selected measure of association between variables (e.g., correlation). Theoretical discussion of how and why this works would be rather lengthy and complicated, and is not necessary for present purposes. PBS Research does, however, have existing computer software to accomplish such analysis. By using studies of pair-wise association in parallel with cluster analysis by variables, additional insights may be gained which go beyond those available by either approach alone.

The final step at this stage of analysis will be a regression of price on individual socioeconomic variables and physician descriptor variables. For socioeconomic variables this will be done with data treated at the county level. The dependent variables will be the average prices over all physicians in a county for a small number of specific procedures. Price data for a number of different procedures can be incorporated and treated simultaneously by constructing the dependent variable to be a weighted index of the average prices for each component procedure by county. For each socioeconomic variable, this analysis will test the hypothesis that price (of physician services in a county) is a function of the socioeconomic variable under consideration, also testing the hypothesis that price is (or is not) significantly correlated with that variable. The linearity of each bivariate relation also will be tested.

Similar analyses will be conducted to test the hypothesis of functional relationship and significant correlation of price with each quantitatively expressed physician descriptor variable. Price data here, however, will be individual physician data; and the analysis in general will be treated at the level of individual physician. The association of qualitative physician descriptors with price level can be tested by the contingency table methods, with price data broken down into discrete class intervals. The contingency table

analysis will not reveal anything concerning the form of a functional relationship, but will at least provide a test of independence or relationship between price and each of the qualitative physician descriptors.

It may be noted that these bivariate studies not only will provide information about the relationship of various individual non-price variables to physicians' prices. As an ancillary benefit, they also will provide information -- even though incomplete, and sometimes only by inference -- on factors that appear to be associated with other questions such as their selections of specialty and practice locality, their decisions to accept or reject Medicare assignment, and their decisions to participate or not participate with Blue Shield.

MULTIVARIATE RELATIONSHIP OF PRICE TO OTHER VARIABLES

Knowledge gained from foregoing stages of the project -- most especially from the series of bivariate studies discussed in the immediately preceding subsection--will be applied to formulate alternative multivariate models and hypotheses regarding determinants of physicians' pricing practices.

The tool to be used here is multiple regression analysis. Initially at least, this will be stepwise multiple regression (with parsimony). The computer program used for this analysis estimates the parameters of multiple linear regression equations in a stepwise manner, regressing first on the most strongly correlated independent variable, then adding the next most strongly correlated independent variable, and so on. Insignificantly correlated independent variables in this process are systematically identified and automatically suppressed or eliminated. Outputs include the mean and standard deviation of each variable, with the convariance matrix and/or correlation matrix printed upon request. For each step an analysis of variance table and multiple correlation are output,

also including coefficients, standardized coefficients, standard errors, F ratios, partial correlations and tolerance, if desired. Other optional outputs include summary tables for F ratios, partial correlations and coefficients; predicted values, residuals and data for each case; scatter plots of predicted and observed values versus selected variable values, residuals versus selected variables, and normal probability plots of residuals.

Input data for these regressions will be at the individual physician level. Price information will be prices or pricing charactertistics for individual physicians, and will be matched with physician information at this level; socioeconomic data will characterize the geographic localities within which the individual physicians practice, and also will be matched with the price information for specific individual doctors.

The price variable (dependent variable) in these regressions may be expressed in a variety of ways. The following are suggested as some of the possible ways of expressing this variable:

(a) Individual physicians' actual average prices for selected procedures;

(b) Each physician's relative average charge for selected individual procedures, defined as

$$\frac{P_i}{\text{mean } P_i}$$

where P_i = the individual physician's average charge for procedure i, and mean P_i = the mean average charge for procedure i for the entire population of physicians included in the analysis;

(c) Each physician's aggregate relative average charge over a combination of several different procedures,

$$\sum_i \frac{P_i}{\text{mean } P_i} ;$$

(d) Each physician's mean relative average charge over a combination of several procedures,

$$\frac{1}{n} \sum_{i} \frac{P_i}{\text{mean } P_i}$$

where there are n different procedures, indexed over i;

(e) Each physician's weighted mean relative average charge over a combination of procedures,

$$\frac{1}{\sum_{i} Q_i} \left(\sum_{i} \frac{Q_i \ P_i}{\text{mean } P_i} \right)$$

where Q_i = number of services the physician reported for procedure i, and all other symbols are as defined above.

The preferred measures from among those suggested above would be (d) or (e). Both of these have the advantage of incorporating information for several different kinds of procedures simultaneously, which is desirable since different procedures commonly exhibit different pricing patterns, and this will smooth out the effects of these divergencies. Both also have the advantage of being able to accomodate missing data due to certain doctors having no reported experience in certain procedures. A surgeon, for example, would not necessarily be expected to be performing obstetrical deliveries; and an obstetrician, conversely, would hardly be expected to be doing tonsillectomies. If both of these procedures were included in the standard mix used for purposes of constructing the above measures, the surgeon and the obstetrician both would have missing data for these procedures they do not perform in their practices. The last suggested measure (e) has a possible further advantage in that it also reflects the doctor's utilization pattern; hence, it more completely represents the relative impact of his overall

pricing pattern on both his own income and the medical services marketplace.

Still another way of representing a physician's relative pricing behavior would be in terms of the percentile position of his average charge for a given procedure in the ordered array of average charges for all physicians reporting this procedure and included in the analysis. Conceptually, this measure also could be extended to apply to a mix of procedures by using the mean or weighted mean of the percentile positions of the doctor's average charges for each of the included procedures. Test problems may be run to see which of these price variables produces more reasonable looking regression results. The price measures likely to be most useful and productive are the weighted percentile and the weighted mean relative average charge. Initial priority will be given to these as alternative ways of expressing the price variables.

At least three classes of models will be constructed and tested: first, models regressing the price variable only on quantitatively expressible physician descriptor variables; second, models regressing the price variable on socioeconomic variables alone; and third, models regressing the price variable on the socioeconomic and physician descriptor variables combined. Other models, perhaps without the parsimony feature, can be constructed and tested using only selected variables suggested by prior stages of analysis, or as expressly desired by SSA. Specific hypotheses to be tested are not suggested here at this time, since these can more properly and appropriately be formulated after having benefit of results from all of the preliminary analyses. In a very real sense, this will be an exercise in systematic discovery, with strong heuristic overtones.

One strategic feature of these analyses that already can be anticipated is that it probably will not be sufficient to deal only with the aggregate pricing characteristics of all physicians taken together. Differences in pricing practices

that fail to appear -- or that are only weakly apparent -- for all doctors in the aggregate may emerge more prominently when one looks only at given individual specialties one at a time -- i.e., when the study is compartmentalized into separate but parallel analyses of sub-populations of physicians, each homogeneous with respect to practice specialty. Similarly, otherwise undetectable differences may begin to emerge when the analyses are applied to homogeneous sub-populations of physicians by geographic area. Therefore, regression analyses will be run with data bases representing only physicians of given selected specialties and/or practicing only in selected geographic areas, as well as the aggregate of all physicians of all specialties throughout the entire state. Analyses also will be run with the data base segregated by major type-service -- i.e., with separate and parallel analyses each addressing a homogeneous product class.

ANALYSIS OF PRICE DISCRIMINATION CLASSES

Further analysis at this point will seek to identify discrete classes into which the physician population may be partitioned on the basis of their empirical pricing practices, and further will seek to identify the salient variables in terms of which this partitioning may be explained. The importance of this stage of study is that it goes beyond the notion of "functional relationship," as described by the multiple regression analysis results, and segregates the physician population into explicit discrete classes which are internally homogeneous with respect to pricing practices and between which there is significant price discrimination. Understanding of the factors accounting for these inter-class discriminations may in the short run support the structuring of reimbursement schedules and/or systems that are more accurately reflective of, and compatible with, actual current practices; and in the longer run, may provide new and deeper insights to help guide future

policy formulation and assess the potential impact of proposed policy changes.

Partitioning of the physician population into discrete price discrimination classes will be accomplished by cluster analysis methods. Cluster analysis partitions a set of subjects into two or more subsets, where the members of any given subset are similar to each other, with respect to the overall effect of a number of attributes or variables used to describe each subject. Each subject is described and characterized by an n-dimensional vector whose elements are the values associated with that subject for each of the n variables or attributes used to describe all subjects. Pair-wise "distances" between all subjects are calculated, where the distance measure may be either a Euclidean distance or a standard statistical measure of similarity, such as correlation or chi-square. Subjects are considered to be more or less similar to each other, depending upon how "close" they are to one another on the basis of this distance measure. The set of subjects then is systematically searched and organized into subsets, or clusters, of subjects that are close to each other in the n-space defined by the n variables used to describe the subjects.

In the computer program available at PBS, clustering begins by finding the closest pair of subjects and combining them to form a cluster. The algorithm continues, joining pairs of subjects, pairs of clusters, or a subject with a cluster, until all subjects are in a single cluster. Thus, beginning with the trivial case in which each individual subject is viewed as a single-member cluster, the population of subjects is progressively organized through an iterative amalgamation process which produces ever fewer numbers of ever larger clusters at each iteration. The process stops when it reaches the other trivial extreme in which all subjects have been joined into one universal custer. The clusters formed are identified at each iteration, with an enumeration of the specific subjects

comprising each cluster. The amalgamation process is defined in such a way as to assure formation of clusters that are mutually exclusive and exhaustive.

The subjects in this application would be individual physicians, and the variables used to describe each physician's pricing pattern would be his average charges for each of several different procedures in a standard mix. The clusters then would identify groups of physicians with similar pricing patterns. To avoid scaling problems that could arise from differences in magnitude of characteristic prices for different procedures, the price data should be standardized, perhaps by transforming to relative average charge measures as defined on page 40.

Price-wise homogeneous groups of physicians identified by the cluster analysis will be further analyzed by discriminant analysis to identify and evaluate the salient variables in terms of which price discrimination between these price homogeneous groups may be explained. The clusters emerging from the foregoing analysis will define the discrimination classes for the discriminant analysis. Interclass discrimination will be analyzed and explained in terms of both socioeconomic and physician descriptor variables. As delineated earlier on page 32, the "socioeconomic" variables include health services supply, demand and activity variables.

At a broad conceptual level, the objective of discriminant analysis is quite simple. On the basis of a set of independent variables, the researcher wishes to classify individuals or objects into two or more mutually exclusive and exhaustive categories or classes. For example, basing the decision on an individual's income, occupation, home ownership, and so forth, one may wish to classify him as either a good or bad credit risk. When the analysis is finished, one then asks how well the procedure succeeded in classifying the individuals, and which variables were most effective in discriminating among the individuals.

The basic technical ideas involved in discriminant analysis can be explained

fairly easily. Essentially, there is a simple scoring system that assigns a score to each individual or object. This score is a weighted average of the individual's numerical values of his independent variables (e.g., age, income, education). On the basis of this score, the individual is assigned to the "most likely" category, from among a predefined set of mutually exclusive and exhaustive categories. Suppose, for example, an individual is 30 years old, has an annual income of $15,000, and has 16 years of formal education. Let b_1, b_2 and b_3 be the weights attached to the independent variables of age, income and education, respectively. This individual's score then would be

$$z = b_1(30) + b_2(15,000) + b_3(16).$$

This value of z could be, for instance, a credit score or rating; and the numerical value of z can be transformed into the probability that the individual is a good credit risk. The b_i are called the discriminant coefficients and are mathematically determined by the analytical procedure. The numerical values and signs of the b_i indicate the relative importance of the associated independent variables in their ability to discriminate among the different classes of individuals.

Mathematically, discriminant analysis is much like a Bayesian regression analysis. Prior knowledge on group membership is updated by sample information (the values of the independent variables for each subject) to give posterior knowledge on group membership. The discriminant coefficients are analogous to regression coefficients. However, instead of explaining variance of the dependent variables as in regression, the object of the discriminant function is to correctly classify as many individuals as possible.

Beyond identifying the significant explanatory variables, results obtained from the discriminant analysis will include measures of the significance of

difference between groups, the relative effectiveness of each variable as a discriminator, and a summary of the relative numbers of subjects correctly classified by the discriminant analysis.

PHYSICIAN PARTICIPATION AND DERIVED INCOME

As noted earlier (page 34), data to be captured and developed for each individual physician will include, for calendar year 1975, total payments received from Blue Shield, total payments allowed by Medicare Part B, relative number of Medicare services on which assignment was accepted, and status as participating or non-participating with Blue Shield. In addition to viewing these as independent variables potentially related to price, ancillary further analyses also will concentrate on each of these separately as a time-related variable and/or as a dependent variable related to other factors. Consonant with the confidentiality and proprietary data provisions incorporated into the contract, no such data will be published or otherwise released for specific individual physicians. Results will be released for publication only after being appropriately aggregated and summarized.

With respect to Blue Shield participation and acceptance of Medicare assignment, the following will be included among specific questions to be addressed:

(a) What have been the Pennsylvania trends in Blue Shield participation and in acceptance of Medicare assignment?

(b) What are some of the general patterns associated with acceptance of assignment--e.g., comparative acceptance rates by type of service, physician specialty, geographic area, etc.?

(c) Is there a correlation between Blue Shield participation and acceptance of assignment under Medicare?

Regarding Blue Shield derived income and Medicare derived income, results to be produced will include descriptive information on variations in such income by county and by specialty. To the extent permitted by availability of appropriate data, these Blue Shield and Medicare derived income data also will be developed on a per-capita basis, relative to the applicable population bases in each county. It is doubtful, however, that meaningful per-capita information on Blue Shield derived income can be developed, since reliable Blue Shield enrollment data do not in general exist on a county-by-county basis.

SUMMARY REVIEW OF PHYSICIAN PRICING MODELS

At the completion of all study phases described up to this point, it may be expected that in addition to a number of substantive findings, a variety of models relating to physicians' prices and pricing practices also will have been developed and tested. Principally, they will be concerned with price trends, price discrimination, the determinants of price levels, and the factors driving and influencing price discrimination. The several models generated to this point will be collectively reviewed and evaluated -- in comparison with one another (where there are alternative models to serve essentially the same purpose), in relation to a priori expectations and objectives for the project, and in light of related findings and developments reported in the literature. The report to be produced at this stage of the project will include a summary and evaluation of the family of models developed in the project, recommendations for future study possibilities, and further model development requirements.

V. ALTERNATIVE CLASSIFICATION SYSTEMS FOR PHYSICIAN REIMBURSEMENT

Findings of the relationship studies completed in prior study phases will be further analyzed and evaluated for the purpose of identifying, rationalizing and selecting alternative locality, specialty or other classifications of physicians that could define peer group structures for reimbursement purposes. Results obtained from the cluster and discriminant analyses should be particularly useful for this purpose. Alternative sets of experimental profiles (UCR screens) will be constructed, based upon the alternative classification strategies. Using a test base representative of the utilization patterns and reported charge levels in a period of actual claims experience, simulations of the claims payment process will be used to determine payout amounts that could be expected with each of the alternative sets of profiles. The alternative classification schemes will be evaluated on the basis of a comparison of their simulated relative payout performances and other impact implications.

GENERAL DISCUSSION OF THE PROBLEM

Numerous methods are used by various medicare and private health insurance carriers to group physicians' "usual" (customary) charges to compute "customary" (prevailing) charges.* For both Medicare and its private business, Pennsylvania Blue Shield groups physicians according to each physician's specialty and charge class as defined by geographic area of practice. Each charge class is composed of

*The PBS private business UCR system uses the terms "usual" and "customary" as counterparts for the "customary" and "prevailing" charges, respectively, in the Medicare Part B system.

numerous political subdivisions (usually counties) which are similar in terms of average level of physician charges, and other socioeconomic characteristics, but which are not necessarily contiguous. Some carriers appropriate different specialty and geographic classifications, and still others ignore specialty or geographic differences altogether. The purpose of the work to be performed in this study phase is to examine, compare and evaluate some of the alternatives for defining prevailing fee categories.

Prerequisite to this analysis is a study of the relationships between various physician classifications and charging levels, as addressed in the preceding phases of study. The factors and characteristics identified in those previous analyses as most highly correlated with physicians' charges, and accounting for the discrimination in their pricing practices, will comprise the initial list of natural grouping parameters to be evaluated in this phase. Classification of profiles by various combinations of physician characteristics will be evaluated with respect to several criteria, both subjective and objective. Geographic categories will be descriptively analyzed in terms of physician distribution and other socioeconomic characteristics. Administrative implications of the various alternatives will be discussed and compared.

Objective testing of the alternative classifications will be in the form of simulated profile building and payout analysis. Even in this part of the study, however, a substantial amount of subjective evaluation will enter into the analysis. Past experience has shown that even when profile building and application procedures are well defined, a number of special problems arise, often involving unusual procedure codes, shortages of data, and circumstances of a doctor's practice. To the extent that it is feasible, an attempt will be made to anticipate such considerations and to compensate for them in the simulation analysis. Subjective judgment also

will influence the decision as to which classification methods merit testing by simulation. The simulation procedure consumes much time and effort, and the number of alternatives which can practically be attempted must, therefore, limit testing to only the most promising alternatives. Assuming that the simulated profiles would be built on a 1975 experience base, a file of claims approved in 1976 then could provide a representative test experience data base for payout simulation purposes in terms of reflecting current charging levels and utilization patterns in sufficient volume to permit meaningful interpretation of simulated payout at fairly fine levels of stratification. Less cumbersome, sampled data bases may also be used to perform more numerous, preliminary simulations.

SIMULATION METHODOLOGY

The general methodology employed by PBS in processing claims under Medicare Part B utilizes several different kinds of primary files. After performing standard checks for provider and beneficiary verifications, an EDP record is produced identifying the particular procedure performed and the provider who performed it, along with other data associated with the claim. After passing a variety of other screens and tests, this record then becomes eligible for processing as a payable claim. To accomplish this, the EDP system searches a series of files and attaches pertinent additional data to the record of services. The following files are accessed for additional data during this process:

(1) The Vendor Master File - data extracted includes the provider's specialty, charge class and practice location.

(2) Customary Charge File - data extracted includes the provider's customary charge for the procedure indicated.

(3) Prevailing Fee File - data extracted includes the "prevailing" charge for the procedure according to the associated charge class and specialty of the provider.

The claim is processed for payment by determining the minimum of the reported charge, customary charge or prevailing charge. Still further processing is required in order to determine any deductible and/or copayment amounts that may be applicable. Analogous files and processing procedures are used in PBS private business claims processing.

This claims processing procedure produces a paid claims file, appropriate portions of which can serve as a test base of actual experience upon which experimental reimbursement mechanisms employing alternative locality/specialty or other classifications may be tried.

In effect, experimental customary and/or prevailing charge files (i.e., profiles) are constructed to encompass the modification being tested. The paid claims test base file then is passed, and determinations are made of the payment amounts that hypothetically would have been allowed under the modified profile structure. These determinations are made by using a computer-based model that emulates the logic of key parts of the actual claim payment process described above, taking into account the actual reported charge for the reported service in each record, but applying the appropriately modified system of customary and prevailing charge screens. These simulated allowed charges then can be compared directly with allowed charges determined according to the conventional method, or according to any of the other alternative schemes being tested.

Suppose, for example, an alternative method is suggested which would eliminate provider specialty and geographic area of practice (charge class) as a classification in determining "prevailing" charges. In order to test the effect of this alternative classification structure, an experimental set of "prevailing" charges

would be developed using all providers' customary charges. The effect of this alternative would be tested by pricing a period of paid claims experience, substituting these simulated prevailing charges for the conventional prevailing charges and calculating a new allowed charge figure. The new allowed charge can then be adjusted, if desired, according to the deductibles and copayments associated with the original claim transaction record. By summing allowed charges and payments over the entire experience base, a comparison can be made of the difference, if any, in allowed charges and payments that may be attributed to the alternative method of developing prevailing charges.

COMPARATIVE EVALUATION OF ALTERNATIVE SCHEMES

The various simulations of alternative profile building procedures will be evaluated with respect to several criteria, among which are the following:

(1) Does the method of classification stratify profiles so that insufficient data for prevailing charge calculation would exist in many categories? This would not only constitute an administrative burden, but extensive manual calculations also dilute the effectiveness of classification.

(2) Does the method of classification have an appreciable effect on total simulated payout, or does it seem to make little difference which method is used? Experience and prior research at PBS suggest that some methods of profile stratification are, in effect, little different from no stratification at all.

(3) How do the alternative methods of classification affect simulated payout among various areas and specialties? A seemingly desirable quality of prevailing charges is that they should in practice

effect a nearly uniform distribution of percent of charges paid, and perhaps they should also generate payout that is somewhat in proportion to the number of doctors or volume of claims submitted for each category. These might be referred to generally as "equity" criteria.

(4) Are some classification schemes apparently redundant in that they generate essentially the same simulated payout performance as do other, less complicated methods?

The evaluation of alternative classification schemes will involve comparative rankings of "most preferable" to "least desirable" to some extent. In many cases, however, assessing the advantages of one method over another will depend heavily upon subjective value judgments. In such instances the study will attempt to present each alternative as objectively as possible and to describe those conditions and assumptions under which particular options might appear advantageous.

ANTICIPATED BENEFITS

In summary, this final phase of study will result in identification, testing and comparative evaluation of some of the more interesting and desirable alternatives for classifying physicians on the basis of specialty, practice locality, or other criteria for reimbursement purposes. The results anticipated here are important particularly because they address a key design issue for physician reimbursement systems and could have almost immediate operational applicability. Further, however, critical analysis and evaluation of the system response variations associated with alternative system configurations also may provide insight and guidance to help formulate and/or evaluate future policy options.

DISCRETIONARY ADMINISTRATIVE PRACTICES

Work described up to this point emphasizes comparison of alternative approaches for defining and designing peer group structures, used as a basis for establishing prevailing fee screens and profiles. In general, these studies address a number of questions and alternative propositions concerning the use of geographic area and practice specialty distinctions as a rational and legitimate basis for certain allowable variations in price.

There are, in addition, a number of other areas of administrative practice in which considerable discretionary latitude may be exercised by carriers, and which accordingly exhibit substantial amounts of variation in actual practice and policy interpretation. Still other variations in administrative practice, while not necessarily ever implemented and/or tested anywhere, frequently have been the subject of speculation by those responsible for policy recommendations and formulation. Together, these include such questions as the frequency with which customary and/or prevailing screens should be updated, minimum sample requirements and other statistical and data quality standards for establishing customary and prevailing fees, the length and relative timeliness of time frames for the data bases used in developing updated profiles, the imposition of arbitrary limitations on fee increases that may be allowed; and may extend to examining the possible consequences of such fundamental changes as the complete elimination of either customary or prevailing screens. SSA has indicated interest in including investigations of some of these further areas of administrative practice among the studies to be conducted at this stage of the project.

While PBS also has an interest in studying these kinds of variations in administrative practice, it must be recognized that these questions are not

within the purview of either Blue Shield's original proposal or SSA's original solicitation for work to be done under Project Areas I and II. As such, PBS feels that it cannot at this point firmly commit itself to undertake additional work of such potentially great magnitude under the existing contract. Supplemental studies in this area will be undertaken only as "targets of opportunity," depending upon a determination by PBS at any point in the project that sufficient time, staff and other resources are available.